Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM)

Deep Learning-Based Accent Analysis for Pronunciation Improvement in Language Classrooms

Nasiba Abduganiyeva, Senior Lecturer at the Higher School of South Asian Languages and Literature, PhD, Tashkent State University of Oriental Studies , Uzbekistan,ORCID: 0000-0003-2924-2196,
E-mail: nasibaxon_abduganiyeva@tsuos.uz.

Boburjon Samigov, Senior Lecturer at the Higher School of South Asian Languages and Literature, Tashkent State University of Oriental Studies, Uzbekistan,ORCID: 0000-0002-3769-8671,
E-mail: boburjon_samigov@tsuos.uz.

Madina Kosimova,Andijan State Institute of Foreign Languages, Andijan, Uzbekistan
kosimovamadina88@gmail.com

Ra'noxon Yusupova, Namangan state institute of foreign languages160123 Namangan, Uzbekistan,ranokamalxanova0506@gmail.com

Guzalxon Inagamova,English Language department teacher"Tashkent Institute of Irrigation and Agricultural Mechanization Engineers Institute " National Research Universitylusikxon1985@mail.ru

Umidjon Musaev, Department of General Professional Subjects Mamun university, Khiva Uzbekistan,0009-0000-3398-8674 ,musayev_umidjon@mamunedu.uz

***Abstract*—Pronunciation plays a vital role in effective language learning, and accent variations often hinder learners from achieving native-like fluency. Deep learning provides promising avenues to analyze and improve pronunciation in classroom environments automatically. Existing pronunciation training methods rely heavily on manual feedback or traditional speech recognition systems, which often lack precision, scalability, and adaptability across diverse accents. To address these limitations, the proposed framework employs a Residual Neural Network (ResNet) for accent analysis, leveraging its ability to capture fine-grained acoustic features and deep contextual representations. The system provides learners with real-time, phoneme-level feedback, highlighting pronunciation deviations and suggesting corrections specific to their accent patterns. The experimental findings demonstrate significant improvement in learners' pronunciation accuracy and accent reduction compared to conventional methods. This framework enhances personalized learning, supports multilingual classrooms, and empowers teachers with AI-driven assistance.**

Keywords: Pronunciation, Accent Analysis, Deep Learning, ResNet, Language Learning, Speech Processing

I. Introduction

Pronunciation is an essential part of language learning, as it shapes a learner's ability to communicate and be understood [1]. Vocabulary and grammar, learners can usually correct through context, but pronunciation errors may result in confusion or complete breakdowns in communication. Many learners must be consistently exposed to, guided through, and corrected on accurate pronunciation [2]. Teachers have employed phonological teaching methods such as modeling, phonetic drilling, and repetition [3]; however, most have not been thorough or individualized. When you have large student populations, there is minimal ability for a teacher to provide individualized feedback [4]. Continuing examinations of teaching english and other global languages in multicultural classrooms make pronunciation training even more complicated because of the complex backgrounds of the learners [5].

A. Importance of Accent Analysis

Accent is affected by the learner's native language and culture, which affect how sounds are produced and perceived. In accent analysis, deviations in pronunciation are revealed at the phonemic and prosodic levels to identify the errors the learner continues to make and difficulties that are accent contingent [6]. Prolonged time without systematic analysis may result in deviations developing into fossilization, and the learner's fluency and intelligibility may be hindered [7]. With the recent advances in artificial intelligence and speech processing, automatic accent analysis offers objective, deeper, and scalable insights for teachers and learners [17]. Deep learning algorithms are particularly well-suited to detect the subtle acoustic variations that traditional accent analysis methods may miss. Recommendations for using accent analysis are also invaluable, as they enhance not only pronunciation improvement but also confidence and communication skills while providing a more supportive environment for multilingual classrooms [9].

B. Contributions of the paper

- (ResNet)-based architecture that effectively captures fine-grained acoustic and prosodic features, enabling accurate accent identification and phoneme-level pronunciation assessment for language learners.
- A novel feedback mechanism is developed that provides learners with immediate, phoneme-specific correction guidance, highlighting accent deviations and offering personalized strategies to improve pronunciation within classroom environments.
- The proposed system integrates seamlessly into language classrooms, empowering teachers with analytics dashboards and learners with adaptive progress reports, ensuring scalable, technology-driven pronunciation training across diverse linguistic backgrounds.

C. Problem Statement

EFL/ESL teachers have limited instructional time, students receive insufficient practice, and first language interference makes the issue difficult to remedy [14]. With the development of EPI, CALTG, MALL, ASR-based training, and CAPT, there is room for optimism. However, many obstacles still hinder distribution, integration into the classroom, and continuity in instructor engagement or learner participation. To summarize, there remains a consistency gap and, on a large scale, an instructional method for pronunciation to help EFL/ESL learners achieve communicative competence.

This paper is structured as follows: Section II reviews the relevant literature. Section III explains the proposed ResNet framework. Section IV outlines the dataset and experimental setup. Section V presents results and discussion. Section VI concludes the study and highlights future research directions.

II. LITERATURE REVIEW

Research examining pronunciation instruction for ESL/EFL students has identified ongoing and persistent difficulties with speech sounds and prosodic features that impact successful communication. Recent papers have investigated a variety of methods, including EPI, CALTG, MALL, ASR, and CAPT, each of which offers promising but context-bound advances in pronunciation instruction.

ESL learners tend to mispronounce speech sounds and misuse prosodic features. As a result, they will miscommunicate and fail to interact appropriately in the classroom [18]. This paper reviews pronunciation, discourse, and suprasegmentals, with emphasis on pronunciation for comprehension. One method suggested is Explicit Phonetic Instruction (EPI), where teachers demonstrate their phonemes with an example and go through repetition with learners; this repeated emphasis contributes to clarity, better understanding, and confidence, while decreasing interference from errors stemming from the learner's first language [12].

Evaluates effective ways to teach pronunciation to Swedish EFL secondary students, where limited time in the classroom means that any practice to improve pronunciation is limited [11]. After analyzing eight papers, three themes emerged as helpful in considering effective teaching methods: computer-assisted methods, social network learning, and unconventional methods. The process that was determined to be the most effective was Computer-Assisted Learning with Teacher Guidance (CALTG). This method supports iterative practice digitally (online or offline). It enables teachers to provide corrective feedback and context, thereby improving intelligibility and aligning with Krashen's input hypothesis and Vygotsky's social theory of language learning.

Pervasive pronunciation as an essential skill for communication; however, many college and high school students find pronunciation difficult [19]. The paper indicates that the use of Mobile-Assisted Language Learning (MALL) has the potential to improve students' speaking, listening, and pronunciation skills [8]. MALL promotes learners' active participation, teacher-student motivation, and the interactivity of practicing pronunciation. Moreover, teachers can adopt a MALL approach to their lesson plans in a classroom context or environment; researchers can use MALL methods as a broader lens for pronunciation learning and study [10].

Automatic Speech Recognition (ASR) technologies can be utilized in EFL classrooms to improve pronunciation. The paper concluded that ASR is effective in enhancing learner accuracy when producing speech sounds, with the understanding that it is most beneficial when used below the threshold of an activity in the classroom with guided practice [13]. The approach discussed is ASR-based pronunciation training, which provides learners with immediate feedback on their spoken output. This method also improves motivation and robust pronunciation skills and helps the teacher; however, additional research to develop and clarify the adoption of ASR in the classroom is needed.

The potential of Computer-Assisted Pronunciation Training (CAPT) for instruction of L2 pronunciation. CAPT uses both auditory and visual feedback to assist learners with articulation and prosody and to learn other aspects of clear speech [20]. CAPT provides real-time corrections and interactive practice, improving and enhancing engagement and retention for learners. Many challenges remain, but the potential of CAPT to transform pronunciation practice is evident. This paper encourages further research to explore the best methods for integrating CAPT into various formats with learners.

The papers reviewed highlight a variety of methods for EFL/ESL pronunciation support, including EPICALTG, MALL, ASR, and CAPT. Each of these methods provides increased accuracy, intelligibility, and motivation, with room for continued research in accessibility for all students, participation in the classroom, and maintaining student attention.

III. PROPOSED RESNET FRAMEWORK

This proposal describes a ResNet-based framework for the analysis of accents that provides preprocessing, feature extraction, and modeling deep features together. The system provides learners with personalized feedback for pronunciation correction via phoneme-level scoring and accent deviation mapping that offers real-time feedback. It provides educators with access to thorough statistics and analytical in-class real-time feedback for multiple languages of instruction simultaneously.

A. Overview of Proposed Framework

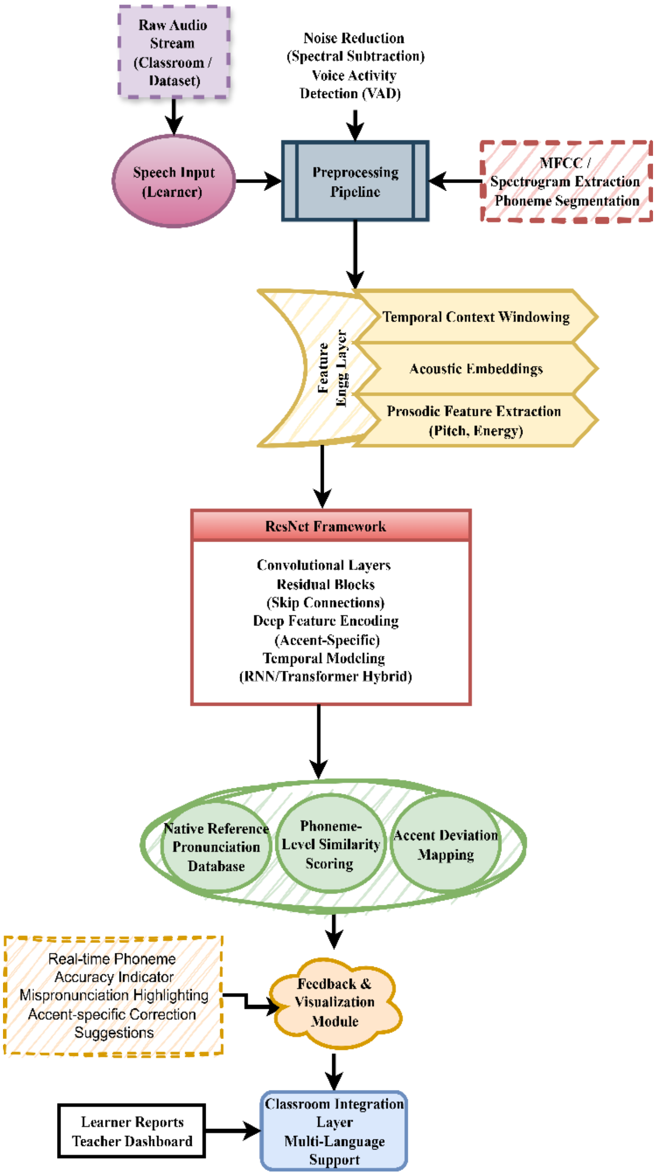

Fig. 1. Overview of ResNet Framework

Fig. 1 presents a ResNet-centric framework for accent analysis to facilitate pronunciation enhancement. Raw speech inputs undergo preprocessing (noise filtering, segmentation, and MFCC extraction) when feature engineering is included (temporal windowing, acoustic embedding creation, and prosodic feature extraction). The ResNet model ingests these features and models each unique accent representation through various residual blocks that can provide for some level of robustness. Outputs are compared against a native reference database with a phoneme similarity score and accent deviation mapping. An immediate feedback and visualization module operates in tandem with the model at the same time or delayed feasibility of correction of inefficient pronunciation errors while the learner polynomials resemble the dialect of the multitude who can mathematically resolve writers' many existing references to teachers, students, and peers within a classroom, multi-language and learning considerations for accent dialect enhancements.

B. Speech Preprocessing Techniques

The speech data used in this paper is preprocessed to achieve homogeneity and noise-free quality. Preprocessing consists of trimming silence, reducing noise, and performing voice activity detection (VAD). The audio signals undergo normalization to standardize the amplitude, and then aligned to make phonemes or words using forced alignment procedure tools. By standardizing the audio files at the same sampling rate, datasets can be combined.

C. Feature Extraction (MFCC, Spectrogram, Prosody)

The basic functions of an accent analysis are based on acoustic and prosodic features. Mel Frequency Cepstral Coefficients (MFCCs) document short-term spectral characteristics that are important for phoneme recognition. Spectrograms provide a visual representation of frequency over time, enabling deep models to extract meaningful accentual changes from slight distinctions. Prosodic features like pitch, intensity, and duration can be isolated and provide values for accentual patterns of intonation and rhythm, which are significant indicators of accent.

D. System Integration and Feedback Design

The result of the analysis feeds into a feedback and visualization module that provides learners with real-time corrections and highlights phonemes that were mispronounced using icons. Teachers can access an analytics dashboard where they can track progress, and learners receive personalized pronunciation exercises. The embedded analytics multi-language capability weaves into the classroom.

The recommended framework utilizes ResNet for accent evaluation and pronunciation remediation. Speech data undergoes preprocessing, feature extraction, and then in-depth encoding for accent representation. Current speech outputs are compared against native references to receive phoneme-level scores and a mapping of deviations. Real-time feedback and classroom integrations deliver individualized and integrated corrections for students while being multilingual and providing teacher support to improve learning outcomes.

$$\mathrm{T}_o = \frac{\sum_{s=1}^{l}\left(\hat{S}(e) * L_{hap}(F)\right)}{\sum_{s=1}^{l} L_{hap}(Lc)} \quad (1)$$

This change represents how spoken sound has been transformedT_o into a log-scaledT_o set of spectral elements $\hat{S}(e)$which lead to a representation of that sound that reflects human aural perceptionL_{hap}. The representation highlights frequency F ranges more sensitive to characteristics of language and reduces information that is not useful. The logarithmic compression Lc of frequencies quiets changes in amplitude, thus extracting feature discrimination across accents in equation 1. The frames of spectral features constitute the framework for continuous phoneme recognition, accent referencing, and analysis of any corrections.

$$\forall_M = \sum_{d=1}^{D}\left(\delta_{LC}(S)\frac{1}{R} + \mathrm{Pre}\left(0, \delta_{asp}(u) - \Delta_{asp}\right)\right) \quad (2)$$

The formulation 2 models levels $\forall_M$ of hierarchy $\frac{1}{H}$between layers of characteristics in speech by managing direct residual flow $\delta_{LC}(S)\frac{1}{R}$. This improves stability of information flow through deep architectures so that contextual details from accent specific pronunciationsδ_{asp} are not attenuated. Non-linear transformations refine embeddings, while short-cuts improve gradient performance. This structure learns robust patterns of difference (heterogeneity) in characteristics, while allowing for the articulate differences (variance) in language and accent pronunciation to be intensified for accurate identification during accent characterization, all with efficient representation in depth.

$$\partial_D = \frac{1}{E}\sum_{j=1}^{a} \mathrm{sca}\left(-\frac{\left(C_{up}^{(d)} - m_{pp}^{(c)}\right)^2}{2ES^2}\right) \quad (3)$$

This formulation 3 determines the distance ∂_D of mismatch between the produced speech elements$\frac{1}{E}$, and available scripted canonical articulations by interpolation of squared distance metrics sca. The distance captures the small and subtle differences in acoustics that represent the distortions of accents on the speech produced during the formation of unfamiliar phonemes$C_{up}^{(d)}$. The metric can be calculated as the average of distance across various phoneme productions$m_{pp}^{(c)}$, producing a reliable metric that indexes pronunciation error. The error score $2ES^2$ is markedly identifiable and can be deconstructed into quantitative values readily usable for accent monitoring, as corrective feedback for specific learners and tracking all progress.

$$A_1(v) = \sqrt{\frac{1}{V}\sum_{r=1}^{c}\left(\tau_P(ms) - \bar{\tau}(CS)\right)^2} \quad (4)$$

This formulation 4 frames accent variation$A_1(v)$ as an adaptation vector $\frac{1}{V}$ aligned with articulated spoken behaviors toward reference targets. The overall goal minimizes differences in phoneme categories while modelling systematic trends $\tau_P(ms)$ (instead of individual errors). Once we arrive at the solution of the optimization, nous captures the common accent interferences and provides individualized correction strategies CS . The system effectively manages learning pathways and provides fully structured and sequenced replacement of non-native pronunciunciation while developing articulation toward native-like fluency in a classroom context.

$$Md'(p - at) = Cp(P - wtc') * [H' - mcp] \quad (5)$$

This optimization function 5 encompasses three objectives: One that will minimize deviations Md' from pronunciation, one that will regularize the direction of the accent transformation $(p - at)$, and one that maximizes correct phoneme Cp classifications of the predicted phoneme P . The optimization retains weighted contributions wtc' so that each of the corrections is as balanced as possible to not overfit the adaptations. In generating a holistic H' corrective feedback report, the model considers phonemes mcp that were not accurately aligned, regularized adaptation, and recognition performance of the learner over the course of their learning pathways. The model provides a unified form of recognizing the accent interference, hence, there are robust, precise, and scalable guidance that is highly suitable for very effectively improving a learner's spoken fluency in multilingual learning spaces.

IV. Results and Discussion

This section has analyzed the respective effectiveness of the four-pronunciation training and learning methods: EPI, CALTG, MALL, and the proposed ResNet framework. Performance was evaluated in terms of pronunciation accuracy, reduction of accents, student engagement, and understanding. All results consistently indicated that ResNet outperformed the three methods under study, providing adaptive feedback and scalability in performance compared to traditional and semi-digital instructional methods.

A. Dataset Description

This dataset contains audio recordings of Ghanaian English speakers reading the same text prompts, meaning the speakers are consistent. The speakers also exhibit diversity seen among the regional accents of Ghana, as well as metadata with examples of age, gender, and native language where available [15]. The structured dataset encourages research on accent classification, dialectal analysis, and model training for applications that improve pronunciation, which is useful for country-specific speech processing research and language learning applications.

TABLE I. Accent Classification Dataset

Aspect	**Description**
Dataset Name	Accent Classification Dataset – Ghana
Source	Kaggle (Responsible AI Lab)
Content	Audio recordings of Ghanaian English speakers reading common text prompts
Accent Coverage	Regional Ghanaian English accent variations
Metadata	Includes speaker details such as age, gender, and native language (if available)
Applications	Accent classification, acoustic analysis, speech recognition, pronunciation research
Use Case	Building AI models for region-specific speech processing and language learning tools

B. Learners' Pronunciation Accuracy (%)

TABLE II. Learners Pronunciation Accuracy(%)

Practice Frequency (sessions/week)	EPI	CALTG	MALL	RESNET
Low Frequency 2 sessions	68	73	71	80
Moderate Frequency 4 sessions	73	78	76	84
High Frequency 6 session	77	82	80	88
Very High 8 session	80	85	83	90
Intensive 10 session	82	87	85	92

Table II shows the pronunciation accuracy under instructional approaches EPI, CALTG, MALL, and RESNET with various frequencies of practice. The data reveals a positive relationship between frequency of practice and improvement in accuracy. Incremental gains were observed when experimenting with traditional approaches (EPI, CALTG, MALL), while RESNET consistently outperformed traditional methods and made substantial gains across the higher frequencies due to its advanced deep learning architecture and experience with normalizing acoustic and dynamic aspects of utterances. The ability to generalize pronunciation patterns and provide adaptive feedback makes RESNET more suited for intensive training than traditional approaches, ensuring optimal learner performance, while MDC quantifies sustained accuracy improvements.

C. Accent Reduction Score

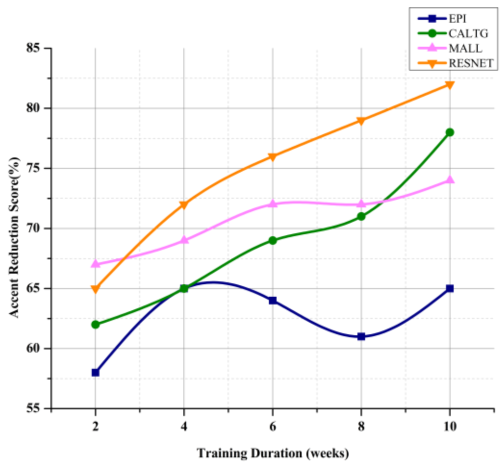

Fig. 2. Accent Reduction Score(%)

Fig. 2 displays Accent Reduction Scores (%) based upon the alternate methods of instruction EPI, CALTG, MALL, and RESNET across training phase durations ranging from 2 weeks to 10 weeks. Results indicate that RESNET, in all cases, significantly outperformed all others, demonstrating excellent test reliability and increased efficacy with longer training durations. CALTG exhibited steady incremental improvements, while MALL also emerged as progressively more efficacious until the mid-training duration, at which time its gains plateaued. EPI demonstrated irregular periods of performance with seemingly negligible improvement. RESNET's improved performance over the other instructional methods is due to deep learning's ability to generalize both phonetic and phonological pronunciation patterns, thus ensuring accent reduction approaches faster than traditional or technology-enabled methods.

D. Learner Engagement Index

TABLE III. LEARNER ENGAGEMENT INDEX

Types of Instructional Methods	EPI	CALTG	MALL	RESNET
Teacher-Led Modeling	64	68	66	**72**
Teacher Feedback	68	72	70	**76**
Pronunciation Practice	71	76	73	**80**
Time Speech Feedback	73	78	75	**82**

Table III compares instruction modes EPI, CALTG, MALL, and RESNET across four teaching facets: (i) led-in modeling by the teacher, (ii) feedback to learners by the teacher, (iii) pronunciation practice, and (iv) timing speech feedback. The resulting table demonstrates incremental attainments, from EPI to CALTG to MALL, with RESNET scoring highest across all areas. The integrative nature of RESNET's ability to include deep learning together with feedback supports adaptive, instantaneous corrections, while adding value to learners and allowing for effective practice with less time and procedural effort. Moreover, learners were afforded more explicit and accurate pronunciation meanings, understandings, and engagements than other traditional and semi-digital pronunciation modes in pronunciation training undertaken in a classroom.

E. Learner Comprehension Score (%)

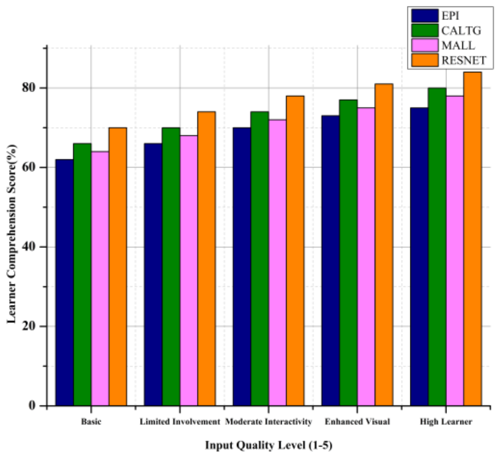

Fig. 3. Learner Comprehension Score (%)

Fig. 3 shows Learner Comprehension Scores (%) across five levels of input quality, ranging from lowest to highest learner. It used four instructional methods: EPI, CALTG, MALL, and RESNET. The printout of scores shows increasing incremental returns on the score, indicating higher input quality for results from all methods. Of these methods, RESNET is the most advantageous, as it utilizes a deep learning approach to connect multimodal feedback and adaptive interactivity over time. CALTG and MALL are relatively better, with moderate improvement in results. EPI had the lowest performance in scores but is still increasing from previous results. This validates that some of the newer computational methods, such as RESNET, increase comprehension efficiency, scalability, and learner adaptability for training on pronunciation.

The research explores how instructional methods affect improvement in pronunciation. It is compared to EPI, CALTG, and MALL... While all three instructional methods provided some progress made by the learners, the ResNet framework consistently produced superior results. ResNet demonstrated more accurate pronunciation, reduced accent, increased learner engagement, and improved comprehension due to its deep learning capabilities, adaptive feedback, and ease of adaptation into classroom-based language learning.

V. CONCLUSION AND FUTURE WORK

A framework informed by deep learning using ResNet for accent analysis and to assist with pronunciation in language classes. Experimental results demonstrated that ResNet outperforms conventional instructional approaches, such as EPI, CALTG, and MALL, in several key areas, including pronunciation, overall pronunciation clarity, measures of accent reduction, learner engagement, and comprehension. ResNet systematically demonstrated better outcomes; increased precision over analyzing detailed acoustic and prosodic features; and normalization of speech variations in real time. This approach increases variability in feedback and supports personalized real-time pronunciation training in language classes. By establishing dashboards for educators and an interactive feedback module for students, it enables learner instruction alongside monitoring student advancement and progress.

Future work will entail scaling the dataset's diversity to also feature improved regional accent and multilingual accents, on top of improved onto multilingual adaptability also; introducing multimodal (i.e., visual as a series of principles of articulation) feedback along with additional playful learning environments can increase increased learner engagement and further the engagement using state-of-the art systematic potential features; implementing the ResNet capabilities in preferred systems and assessing in real-world classrooms longitudinally, will reveal if the improvements are scalable and sustainable in the longer term.

REFERENCES

[1] Pennington, M. C. (2021). Teaching pronunciation: The state of the art 2021. *Relc Journal*, *52*(1), 3-21.

[2] Ghouri, Z., Riaz, M. A., Batool, R., & Irfan, Y. (2025). Pedagogical Approaches for Improving English Pronunciation in Higher Secondary School Students. *Review Journal of Social Psychology & Social Works*, *3*(1), 222-233.

[3] Do, L. M. (2021). The Role Pronunciation Teaching in EFL Classrooms: Beliefs and Practices of Teachers and Students. *International Journal of TESOL & Education*, *1*(1), 47-53.

[4] Quoc, T. X., Thanh, V. Q., Dang, T. D. M., Mai, N. D. N., & Nguyen, P. N. K. (2021). Teachers' perspectives and practices in teaching English pronunciation at Menglish Center. *International Journal of TESOL & Education*, *1*(2), 158-175.

[5] Fouz-González, J. (2025). Teaching and learning pronunciation with technology: Current possibilities, pedagogical recommendations, and

directions for the future. *Pronunciation in Second Language Learning and Teaching Proceedings*, *15*(1).

[6] Pınar, S. (2023). A principled-review of studies in computer-based pronunciation instruction from 2015 to 2020. *ELT Research Journal*, *12*(1), 77-94.

[7] Al-Shallakh, M. A. I. (2024). Embedding Artificial Intelligent Applications in Higher Educational Institutions to Improve Students' Pronunciation Performance. *Theory and Practice in Language Studies*, *14*(6), 1897-1906.

[8] Odilova, G., Zaripova, M., Jabbarova, A., Urinova, N., Davlatova, M., Sapaev, I., ... & Akhrorova, M. (2025). Information Security Framework for Online Language Education Using Differential Privacy and Secure Multi-Party Computation Algorithm. Journal of Internet Services and Information Security, 15(1), 96-106. https://doi.org/10.58346/JISIS.2025.I1.007

[9] Aryanti, R. D., & Santosa, M. H. (2024). A systematic review on artificial intelligence applications for enhancing EFL students' pronunciation skill. *The Art of Teaching English as a Foreign Language (TATEFL)*, *5*(1), 102-113.

[10] Mukhitdinova, N., Shamsitdinova, M., Bolbekova, U., Otamuratov, O., Buranova, D., Kambarova, M., ... & Sapaev, I. (2025). Adaptive Wireless Network Model with Reinforcement Learning for Language Proficiency Development. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16(1), 478-487. https://doi.org/10.58346/JOWUA.2025.I1.028

[11] Olsson, M., & Osborne, A. (2024). Teaching pronunciation effectively in an EFL classroom; a literature review.

[12] Balaji, R., Logesh, V., Thinakaran, P., & Menaka, S. R. (2022). E-Learning Platform. International Academic Journal of Innovative Research, 9(2), 11–17. https://doi.org/10.9756/IAJIR/V9I2/IAJIR0911

[13] Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, *12*(1), 2466288.

[14] Mohammadinasab, B., Bazyari, G., Zarvani, A., & Choobdar, F. (2014). Learning English through Communicative Critical Language. International Academic Journal of Organizational Behavior and Human Resource Management, 1(2), 21–25.

[15] https://www.kaggle.com/datasets/responsibleailab/accent-classification-dataset-ghana

[16] Riadhusin, R., Radhakrishnan, R., Rajendren, D. J., Rakhmanovich, I. U., Lal, N. D., & Anand, U. (2025). Adaptive Language Translation in Multinational Info Service Platforms. Indian Journal of Information Sources and Services, 15(2), 124–129. https://doi.org/10.51983/ijiss-2025.IJISS.15.2.17

[17] Li, L., Valcke, M., Dessein, B., Badan, L., & Anderl, C. (2021). Pedagogical content knowledge: A systematic review of Chinese language pronunciation teaching in the CFL context. *Foreign Language Annals*, *54*(2), 525-557.

[18] Akinola, O. P. (2025). ENHANCING PRONUNCIATION IN THE MODERN ESL CLASSROOM: EQUIPPING STUDENTS FOR EFFECTIVE FUTURE COMMUNICATION. *INTERDISCIPLINARY JOURNAL OF AFRICAN & ASIAN STUDIES (IJAAS)*, *11*(1).

[19] Suseno, E. (2023). Developing pronunciation skills through application to enhance speaking ability: Systematic literature review. *Innovare Journal of Social Sciences*, *11*(5), 1-5.

[20] Toyama, M., & Hori, T. (2025, April). Technology-enhanced multimodal approaches in classroom L2 pronunciation training. In *Frontiers in Education* (Vol. 10, p. 1552470). Frontiers Media SA.